



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Words and the World

**Citation for published version:**

Lupyan, G & Clark, A 2015, 'Words and the World: Predictive Coding and the Language-Perception-Cognition Interface', *Current Directions in Psychological Science*, vol. 24, no. 4, pp. 279-284.  
<https://doi.org/10.1177/0963721415570732>

**Digital Object Identifier (DOI):**

[10.1177/0963721415570732](https://doi.org/10.1177/0963721415570732)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Current Directions in Psychological Science

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



**Words and the World: Predictive coding and the language-perception-cognition interface.**

Gary Lupyan  
University of Wisconsin-Madison  
[lupyan@wisc.edu](mailto:lupyan@wisc.edu)

Andy Clark  
Edinburgh University  
[andy.clark@ed.ac.uk](mailto:andy.clark@ed.ac.uk)

<sup>1</sup>Gary Lupyan  
Department of Psychology  
1202 W. Johnson St.  
Madison, WI 53703

## Abstract

Can what we know change what we see? Does language affect cognition and perception? The last few years have seen increased attention to these seemingly disparate questions, but with little theoretical advance. We argue that substantial clarity can be gained by considering these questions through the lens of ‘predictive processing’, a framework according to which mental representations—from the perceptual to the cognitive—reflect an interplay between downward-flowing predictions and upward-flowing sensory signals. This framework provides a parsimonious account of how (and when) what we know ought to change what we see, and helps us understand how a putatively high-level trait such as language can impact putatively low-level processes like perception. Indeed, within predictive processing, language begins to take on a surprisingly central role in cognition by providing a uniquely focused and flexible means of constructing predictions against which sensory signals are evaluated. Predictive processing thus provides a plausible mechanism for many of the reported effects of language on perception, thought and action, and new insights on how and when speakers of different languages construct the same ‘reality’ in alternate ways.

## 1. The Predictive Brain

Across the cognitive sciences, a picture is emerging in which the brain is viewed as an engine of probabilistic prediction. On this view, every level in the hierarchically organized system that comprises the brain attempts to predict the activity in the level below it (Figure 1). A remarkable consequence of this arrangement is that seeking to reduce the overall prediction error produces representations at multiple levels of abstraction, flexibly incorporating whatever sources of knowledge help to reduce the overall prediction error. The higher-level (more abstract) representations formed in the goal of minimizing prediction error in the present, enable better predictions in the future.<sup>1</sup>

We begin with a brief outline of this framework and then apply it to two domains which, on the surface, seem to have little to do with each other, but are unified under the new framework: the

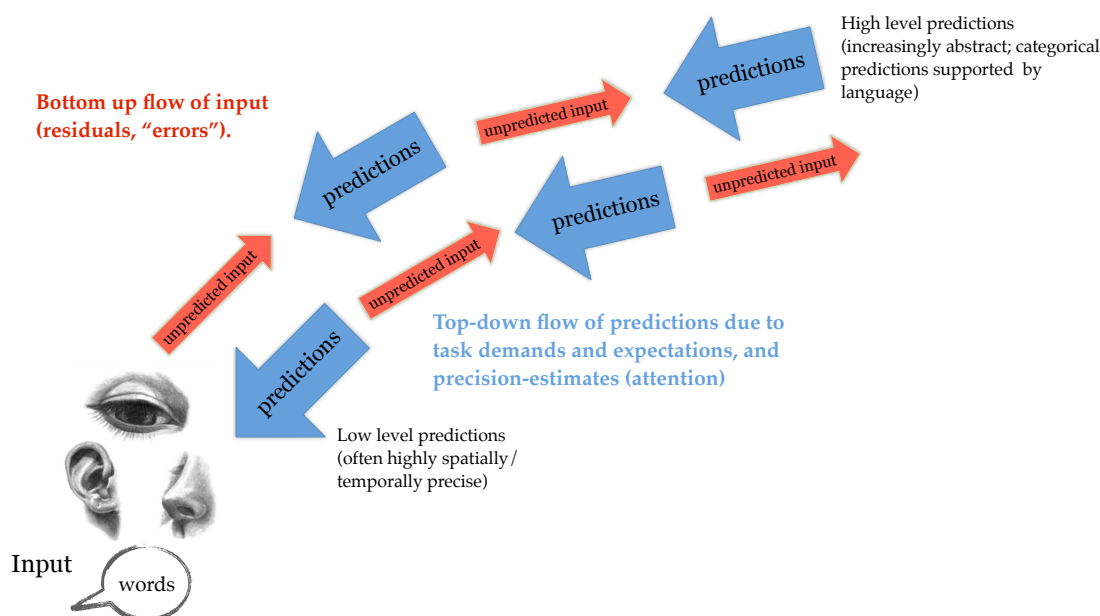


Figure 1. A highly schematized view of the predictive processing view of information transfer in the brain. Bottom-up inputs are processed in the context of priors (beliefs/hypotheses) from levels higher-up in the hierarchy. The unpredicted parts of the input (errors) travel up the hierarchy, leading to the updating of subsequent predictions, and the cycle continues. The relative contribution of the bottom-up signal is determined by varying precision estimates. A highly variable or imprecise signal is given less weight.

cognitive penetrability of perception, and effects of language on perception, action, and ‘thought’ more broadly.

Consider the image in Figure 2 (left)—the so-called ‘Cornsweet Illusion’. To most people, the central paired tiles appear to be very different shades of grey—an appearance that, as the second picture reveals, is illusory. The illusion occurs because our visual experiences do not veridically reflect the current inputs, but are informed by ‘priors’ (prior beliefs, usually taking the form of non-conscious predictions or expectations) concerning the world. In this case, the prior is that surfaces tend to be equally reflectant rather than becoming gradually brighter or darker towards their own edges. The brain’s best guess is that the central pairing involves two differently reflective surfaces (two different shades of grey) illuminated by differing amounts of light. The illusion thus occurs because the image displays a highly atypical combination of illuminance and reflectance properties and the brain uses what it has learnt about typical patterns of illumination and reflectance to infer (falsely in this case) that the two tiles must be different shades of grey. In the world we actually live

in, these particular prior beliefs or neural expectations are provably ‘Bayes optimal’ – that is, they represents the globally best method for inferring the state of the world from the ambient sensory evidence (Brown & Friston, 2012).

The brain, on these accounts, combines prior knowledge or expectations (including knowledge about the present context) with the incoming sensory evidence to yield a percept that reflects its best available hypothesis concerning the most probable state of the world. This view of ‘perception-as-inference’ originates with Helmholtz (1860) and has had many more recent champions, including Ulric Neisser and Richard Gregory. It is only in recent years, however, that these broad visions have been given effective computational flesh, shown to be (roughly speaking) neurally plausible, and seen to converge with compelling bodies of work in psychophysics and cognitive psychological showing that much of perception conforms to optimal (Bayesian) ways of combining sensory evidence with prior knowledge.

*Predictive processing* (a term we use for models that implement hierarchical predictive coding) is a framework that combines all these elements (Clark, 2013; Friston, 2010; Hohwy, 2013). Predictive processing approaches share many features with earlier work in connectionism/parallel distributed processing, such as McClelland & Rumelhart’s Interactive Activation model (for some recent discussion, see McClelland, 2013; McClelland, Mirman, Bolger, & Khaitan, 2014), but they add an important emphasis upon hierarchical structure, and the ongoing attempt to predict the current sensory signals (see also Hinton, 2007). A key emphasis of predictive processing models is an asymmetry between the forward and backward flow of information: the forward flow computes residual errors, while the backwards flow delivers predictions. Within the predictive processing framework, percepts emerge via a recurrent cascade of ‘top-down’ predictions that involve expectations spanning multiple spatial and temporal scales. The downward predictions reflect what the system expects given what it already ‘knows’ about the world and about the current context. These predictions are combined with incoming sensory data to arrive at progressively better guesses about the source of the signal (the world). Aspects of the input that are unexplained are sent forward as ‘prediction error signals’ which ‘carry the news’, by pushing unexplained elements of the sensory signal upwards so as to select new top-down hypotheses that are better able to accommodate the present sensory signal. This process runs concurrently and continuously across multiple levels of a processing hierarchy. While most of the predictions are unconscious, one can sometimes become aware of them when they are violated. For example, imagine drinking from a glass of what you think is orange juice only to realize on tasting it that it is actually milk. The difference between the experience of tasting milk when

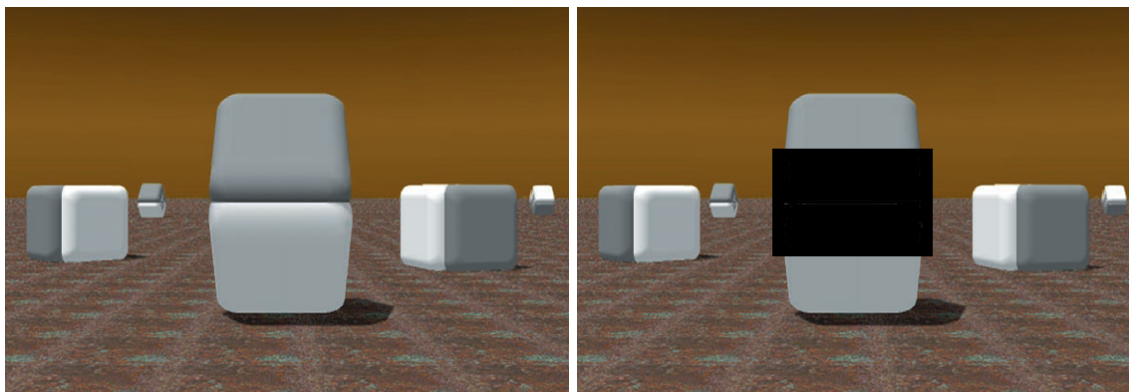


Figure 2: The first image depicts a typical Cornsweet illusion set up. The centre of the two tiles comprising the central pairing appear to be different shades of grey. The second image reveals that they are in fact the same shade of grey. Image source and credit: Purves, D., Shimp, A., & Lotto, R. B. (1999). An empirical explanation of the Cornsweet effect. *The Journal of Neuroscience*, 19(19), 8542-8551.

expecting it, and when expecting orange juice instead is an expectation made tangible (Lupyan, 2015). Similarly, consider the experiential impact of an unexpected omission, as when a musical note is missed out of a familiar composition. Such omissions can be as perceptually striking and salient as the most vibrant tone – an otherwise puzzling effect that is neatly explained by assuming that the construction of perceptual experience involves expectations based upon some kind of model of what is likely to occur.

The perceptual problems that confront us in daily life vary greatly in their demands. For some tasks, it is best to deploy large amounts of prior knowledge, while for others it may be better to let the world do as much of the driving as possible. Which strategy is best is also hostage to myriad contextual effects. Walking around our own house in the dark, it may be wise to let detailed top-down knowledge play a substantial role. Driving fast along an unfamiliar winding mountain road, we need to let sensory input take the lead. How is a probabilistic prediction machine to cope? It copes by continuously estimating and re-estimating its own *sensory uncertainty*, assigning more or less weight to top-down expectations versus bottom-up inputs accordingly in the service of minimizing overall prediction error. Within this framework, estimations of sensory uncertainty modify the impact of prediction error signals at each level of processing according to their estimated ‘precision’, where this is the brain’s best guess at their certainty or reliability (inverse variance, for the statistically-savvy). Variable precision-weighting is thus a mechanism for tuning the extent to which input is modulated by top-down predictions. As we shall see, this mechanism also provides a way for language to serve as a superbly flexible tool for transforming sensory processing.

## 2. Perception as a predictive and penetrable process.

Viewing perception as a predictive process helps to resolve a longstanding argument concerning whether perception is “penetrated” by knowledge (Pylyshyn, 1999; Lupyan, 2015). Within the predictive processing framework perception is expected to be penetrable to the extent that such penetration minimizes overall (long-term) prediction error. If information from another modality, prior experience, expectations, knowledge, beliefs, etc., lowers overall prediction error, then this information will be used to guide processing at these lower levels (we reiterate that this process is not a “decision” but the consequence of minimization of the prediction error<sup>1</sup>). In some cases this penetration changes what we consciously experience (e.g., the lightness of the tiles in Figure 2). In other cases, the conflict between bottom-up inputs and top-down predictions can be resolved at higher level. For example in Figure 3 the meaning of the central image changes depending on its context, but what we literally see is (relatively) unaffected (at least when are free to examine it in detail).

There is, of course, no gatekeeper deciding the extent to which a cognitive state should penetrate perception. In contexts where altering the activation patterns at lower levels minimizes overall prediction error, we should find that what we know changes what we see. In other situations lower-level processing is untouched and conflicts between predictions and inputs are resolved at higher-up levels that are often referred to as decisional



Figure 3. Another example in which the prior local contextual cues set up expectations. In the context of reading the A, the B hypothesis makes the raw visual data most probable. In the context of reading the 12, the 13 hypothesis makes the very same raw visual data most probable. Unlike the Cornsweet illusion, the central image remains visually ambiguous even as its meaning is disambiguated because the top-down prediction can be integrated with the bottom-up signal at a relatively higher level.

or post-perceptual, corresponding to situations in which our phenomenology is relatively unaffected by knowledge and expectations. The predictive-processing framework offers a precise way to strike this balance according to the estimated reliability (more on this below) of the prediction error signal at different levels of processing. Contextual information thus enables the system to allow expectations to penetrate perceptual processing more deeply in some contexts than in others depending on the reliability of the input.

This framework helps to resolve several persistent confusions: (1) It is commonly argued that if what we knew changed what we saw, then knowing that the two tiles in Figure 2 are actually the same lightness ought to cause us to see them that way (e.g., Pylyshyn, 1999). The problem is that discounting the input is incompatible with long-term error reduction. If the illusory percept offers the best prediction in the majority of situations, then, in the long term, the illusion is Bayes optimal. Simply letting a belief override a bottom-up input will, in many cases, result in very high prediction error; the input and the higher-level belief need to be weighted according to their respective likelihoods. (2) Critics of cognitive penetrability contend that many demonstrations of effects of beliefs/knowledge/expectations on perception are merely attentional such that knowledge can affect what one attends to, but not how the attended inputs are subsequently processed (Lupyan, 2015, for review). In contrast, within the predictive processing framework, attention is not something one “focuses” or “deploys” (Anderson, 2011). Rather, it is the mechanism of variable precision-weighting itself. When one “attends” to something, small deviations from expectations are weighed more than when one is not attending to it (den Ouden, Kok, & de Lange, 2012; Feldman & Friston, 2010). As a result, the neural representations of objects that are being attended (because they are task-relevant) are measurably different than the same object when it is not attended (e.g., Çukur, Nishimoto, Huth, & Gallant, 2013).

### 3. Predictive processing and the relationship between language, perception, and ‘thought’

A commonly held view is that the sole function of language is to communicate our thoughts. On this view, words and larger linguistic constructions latch onto pre-existing concepts, enabling highly flexible communication, but without altering the workings of ‘nonverbal’ systems involved in, e.g., categorization, memory, and perception (Pinker, 1994; Snedeker & Gleitman, 2004). A corollary of this view is that although different languages provide their speakers with different ways of *talking* about things (Malt et al., in press), these differences have nothing to with how we *think* or *perceive* things (Gleitman & Papafragou, 2005).

A flurry of findings from cognitive and developmental psychology, however, argue for a much more transformative role of language both in higher-level cognitive processes and in basic perception (Boroditsky, 2010; Casasanto, 2008; Lupyan, 2012 for reviews): rather than passively reflecting the joints of nature, words and larger constructions help carve joints *into* nature. Language functions *not only* as a means of communicating our thoughts, but plays an active role in shaping them.

To cite but two examples: controlled studies of the famous ‘Eskimo words for snow’ thought experiment show that under certain conditions having labels indeed facilitates learning new categories (Lupyan, Rakison, & McClelland, 2007). Once learned, verbal labels appears to be uniquely effective in activating conceptual content (Lupyan & Thompson-Schill, 2012). Considering the relationship between language and thought within the framework of predictive processing allows us to go beyond these individual observations toward a fuller more unifying account.

#### 3.1 Words as artificial contexts

We take for granted our ability to change people’s behavior using language. It is tempting to think of the process by which we understand a phrase like “attend to the vehicles” as activating a

repository of stored knowledge into which incoming perceptual input is then slotted and appropriate actions selected. A very different perspective is that linguistics input affects mental states just like other perceptual inputs (Elman, 2009). Consider the study of Çukur and colleagues (2013): participants undergoing fMRI watched movie clips passively or while monitoring for humans or for vehicles while. These prompts shifted neural representations throughout the brain (including primary visual cortex) such that a prompt to attend to vehicles expanded the neural representations of vehicles and semantically related entities, while collapsing semantically distant categories. Such findings are consistent with claims from “embodied cognition” that words activates neural patterns overlapping with those activated by actual vehicles (Lupyan & Bergen, in press).

However, words are very specific kinds of cues. Any direct experience of a vehicle is an experience of a *specific* vehicle, the word “vehicle” is *categorical*.<sup>2</sup> Moreover, we can flexibly modify “vehicle” by referring to a “small vehicle” or “an upside-down vehicle” (this is what grammar is for, after all). Verbal cues of this sort (even if self-generated) can act as flexible *categorical* inputs allowing an organism to weigh incoming input possibly through a broad, fast re-tuning of the organism’s entire semantic network of the sort shown by Çukur et al. (2013). If this is true, we might expect that simply hearing a word can lead the visual system to generate a predictive signal helping to process an input that is otherwise too weak or noisy. Indeed, in a recent study, simply hearing a word boosted otherwise invisible images into awareness (Lupyan & Ward, 2013).

Viewed from the perspective of predictive processing, words are seen not just tools for communication, but as highly flexible (and metabolically cheap) sources of ‘artificial contexts’. Language directed at others and at oneself (e.g., in verbal rehearsal and other forms of self-directed speech) provides a powerful tool for manipulating thought and reasoning. The main mechanism by which language accomplishes this manipulation is through flexible modification of both what top-down information is brought to bear, and (by selectively influencing the precision-weighting of prediction error) how much influence this top-down information has on specific lower-level level processes. Language can thus help constrain what representations are recruited, and what impact they have on reasoning and inference. On this view, language becomes a powerful tool for cognitive self-manipulation, providing a huge boost to intelligence.<sup>3</sup> The idea that words and larger constructions thus afford a kind of flexible ‘programming language’ for the mind would help explain why language is so persistently linked to such a wide range of behaviors. For example, vocabulary size and other verbal measures are surprisingly good predictors of performance on “nonverbal” intelligence tests like Raven’s Matrices (e.g., Cunningham & Stanovich, 1997) while linguistic impairments are linked to marked deficits (Baldo, Bunge, Wilson, & Dronkers, 2010).

In sum, we propose that the learning of language may create a potent means of biasing the recruitment of prior knowledge and of artificially manipulating, at any level of processing, the weightings that determine the relative influence of different top-down expectations and incoming sensory signals. Such transient, targeted, manipulations could selectively enhance or mute the influence of any aspect, however subtle or complex, of our own or another agent’s world model. Exposure to language (whether shared or self-produced) would then emerge as a potent and fundamentally unified means of exploring and exploiting the full potential of our own acquired knowledge about the world - a kind of artificial ‘second system’ enabling us to make fully flexible use of what we know.

#### 4. Further reading

Clark (2013) and Hohwy (2013) explore hierarchical predictive coding (“predictive processing”) as a unifying framework for understanding neural processing in article and book form, respectively. Frith and Frith (2012) apply principles of predictive coding to social cognition and mentalizing. Friston (e.g., Feldman & Friston, 2010; Friston, 2010) summarizes computational accounts of hierarchical



predictive coding and its relationship to competing theories. Lupyan (to appear, 2015) reviews the evidence for/against cognitive penetration of perception, focusing on predictive coding; Lupyan (2012) reviews evidence of effects of verbal labels on cognition and perception within a theory of 'language augmented thought.'

## 5. References

- Anderson, B. (2011). There is no such thing as attention. *Frontiers in Theoretical and Philosophical Psychology*, 2, 246. doi:10.3389/fpsyg.2011.00246
- Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain and Language*, 113(2), 59–64. doi:10.1016/j.bandl.2010.01.004
- Boroditsky, L. (2010). How the languages we speak shape the ways we think: the FAQs. In M. J. Spivey, M. Joannisse, & K. McRae (Eds.), *The Cambridge Handbook of Psycholinguistics* (p. forthcoming). Cambridge: Cambridge University Press.
- Brown, H., & Friston, K. J. (2012). Free-Energy and Illusions: The Cornsweet Effect. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00043
- Casasanto, D. (2008). Who's afraid of the big bad whorf? Crosslinguistic differences in temporal language and thought. *Language Learning*, 58(1), 63–79.
- Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences*, 36(3), 181–204.
- Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, 16(6), 763–770. doi:10.1038/nn.3381
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33(6), 934.
- Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, 3, 548. doi:10.3389/fpsyg.2012.00548
- Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33(4), 547–582. doi:10.1111/j.1551-6709.2009.01023.x
- Feldman, H., & Friston, K. J. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, 4. doi:10.3389/fnhum.2010.00215
- Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. doi:10.1038/nrn2787
- Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159(3), 417–458.
- Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, 63, 287–313. doi:10.1146/annurev-psych-120710-100449
- Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.), *Cambridge Handbook of thinking and Reasoning* (pp. 633–661). Cambridge: Cambridge University Press.
- Helmholtz, H. von. (1860). *Handbuch der physiologischen optik*. (J. P. C. Southall, Ed.) (Eng Trans., Vol. 3). New York: Dover.
- Hinton, G. E. (2007). Learning multiple layers of representation. *Trends in Cognitive Sciences*, 11(10), 428–434. doi:10.1016/j.tics.2007.09.004
- Hohwy, J. (2013). *The Predictive Mind*. New York, N.Y.: Oxford University Press.
- Lupyan, G. (2012). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 57, pp. 255–297). Academic Press. Retrieved from <http://www.sciencedirect.com/science/article/pii/B9780123942937000078>

- Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*.
- Lupyan, G., & Bergen, B. (in press). How language programs the mind. *Topics in Cognitive Science*.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1082.
- Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology-General*, 141(1), 170–186. doi:10.1037/a0024904
- Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, 110(35), 14196–14201. doi:10.1073/pnas.1303312110
- Malt, B. C., Gennari, S. P., Imai, M., Ameel, E., Saji, N., & Majid, A. (in press). Where are the Concepts? What Words Can and Can't Reveal. In E. Margolis & S. Laurence (Eds.), *Concepts: New Directions*. Cambridge: MIT Press.
- McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Frontiers in Psychology*, 4. doi:10.3389/fpsyg.2013.00503
- McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive Activation and Mutual Constraint Satisfaction in Perception and Cognition. *Cognitive Science*, 38(6), 1139–1189. doi:10.1111/cogs.12146
- Pinker, S. (1994). *The Language Instinct*. New York: Harper Collins.
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3), 341–365.
- Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *Weaving a Lexicon* (illustrated edition., pp. 257–294). Cambridge, MA.: The MIT Press.

### Notes

1. To be clear, the brain is no more “trying” to predict than a gas at a higher pressure tries to diffuse to a lower pressure. This increase in entropy characterizes all nonliving systems. But living systems can temporarily resist this by avoiding some environments and altering others. Organisms that can predict their own sensory inputs at multiple spatial and temporal scales are well-placed to do this, thus maintaining themselves within their species-specific window of viability (Friston & Stephan, 2007).
2. This claim is not limited to a superordinate term like ‘vehicle’ and holds at any level of abstraction. Any experience with a dog, the color red, an instance of on-ness, or a glimpse of your brother Bob are all *particulars*. The corresponding terms abstract over these particulars in a way that perceptual experience does not.
3. The role of language in intelligence is taken for granted in much of classical and contemporary philosophy of mind, though without much elaboration of mechanism. In contrast, as noted above, major strands of contemporary cognitive science and developmental psychology tend to dismiss language as purely a communicative tool.